Naval Research Laboratory

Washington, DC 20375-5320

NRL/MR/5540--06-8945

The Capacity of a Binary Timing Channel with Noise

KEYE MARTIN

Center for High Assurance Computer Systems
Information Technology Division

March 24, 2006

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
24-03-2006

2. REPORT TYPE
Memorandum Report

3. DATES COVERED *(From - To)*
21-27 September 2005

4. TITLE AND SUBTITLE

The Capacity of a Binary Timing Channel with Noise

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
602235N

6. AUTHOR(S)

Keye Martin

5d. PROJECT NUMBER
NRL/ITD/5540/RD/06/673

5e. TASK NUMBER

5f. WORK UNIT NUMBER
55-6326-06

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Research Laboratory
Center for High Assurance Computer Systems, Code 5543
4555 Overlook Avenue, SW
Washington, DC 20375-5320

8. PERFORMING ORGANIZATION REPORT NUMBER

NRL/MR/5540--06-8945

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

National Research Council

10. SPONSOR / MONITOR'S ACRONYM(S)

11. SPONSOR / MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

We obtain a formula for the capacity of a binary timing channel with general noise in terms of the unique solution of a channel dependent equation. We then give three provably correct algorithms that can be used to solve the equation.

15. SUBJECT TERMS
Information theory
Capacity
Time
Noise

16. SECURITY CLASSIFICATION OF:
a. REPORT Unclassified
b. ABSTRACT Unclassified
c. THIS PAGE Unclassified

17. LIMITATION OF ABSTRACT
UL

18. NUMBER OF PAGES
8

19a. NAME OF RESPONSIBLE PERSON
Keye Martin

19b. TELEPHONE NUMBER *(include area code)*
(202) 404-4909

The capacity of a binary timing channel with noise

Keye Martin

Center for High Assurance Computer Systems (Code 5540)
Naval Research Laboratory
Washington D.C. 20375
kmartin@itd.nrl.navy.mil

Abstract

We obtain a formula for the capacity of a binary timing channel with general noise in terms of the unique solution of a channel dependent equation. We then give three provably correct algorithms that can be used to solve this equation.

1 Introduction

Shannon expressed the capacity of a discrete noiseless channel with variable symbol time durations as the logarithm of the solution of a certain equation. In [1], this analysis was carried out for a certain type of binary channel called a "Z channel" – a channel in which one symbol is subjected to noise but the other is not. As far as we are aware, the timed Z channel of [1] is the only case in which the capacity of a timed binary channel *with noise* has been studied in the literature. In this note, we study the capacity of a binary timed channel with general noise, deriving a formula for its capacity in terms of the unique solution of an equation that depends on the channel. The equation is nonlinear so we give three different numerical methods that can be used to solve it and prove that each of them converge.

2 A formula for the capacity

Consider a channel through which one of two symbols $\{\circ_1, \circ_2\}$ may be sent. The output received is a symbol in $\{\bullet_1, \bullet_2\}$. If $\bullet_1$ is received, the transmission took t_1 units of time; if $\bullet_2$ is received, it took t_2 units of time. The output probabilities $\bar{y} = (y_1, y_2)$ are the product of the input probabilities $\bar{x} = (x_1, x_2)$ with the matrix of conditional probabilities:

$$\bar{y} = \bar{x} \cdot \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

Here $a = P(\bullet_1|\circ_1), b = P(\bullet_2|\circ_1), c = P(\bullet_1|\circ_2), d = P(\bullet_2|\circ_2)$, so $y_1 = (a-c)x_1 + c$ and $y_2 = 1 - y_1$. To calculate the capacity we want to maximize

$$I_t = \frac{H(Y) - H(Y|X)}{E(T)}$$

Manuscript approved January 23, 2006.

over all possible $\bar{x}$. This means we want to maximize the function $I_t : [0,1] \to \mathbb{R}$ given by

$$I_t(x) = \frac{H(f(x)) - xH(a) - (1-x)H(c)}{t_1 f(x) + (1-f(x))t_2}$$

where $H : [0,1] \to \mathbb{R}$ is $H(x) = -x\ln x - (1-x)\ln(1-x)$ and $f : [0,1] \to \mathbb{R}$ is $f(x) = (a-c)x + c$. Notice that $a + b = c + d = 1$. The times t_1 and t_2 are positive real numbers. Recall that entropy H is *strictly convex*: for $x, y \in [0,1]$ and $p \in [0,1]$, we have

$$H(px + (1-p)y) \geq pH(x) + (1-p)H(y)$$

with equality if and only if $p = 0$, $p = 1$ or $x = y$.

Lemma 2.1 *The absolute maximum value of I_t is always assumed at a point in* $(0,1)$.

Proof. If $a = c$, then $I_t \equiv 0$, and the claim is trivially true. Suppose now that a and c are different. By strict convexity of H, we can write

$$I_t(x) > \frac{H(f(x)) - H(xa + (1-x)c)}{t_1 f(x) + (1-f(x))t_2}$$

when $x \in (0,1)$. Then $I_t(1/2) > 0$. However, $I_t(0) = 0 = I_t(1)$, so the maximum value of I_t must be assumed at some point of $(0,1)$. □

Since the maximum of I_t is assumed at a point in the interior of $[0,1]$, the equation $\dot{I}_t = 0$ has at least one solution.

Theorem 2.2 *If a and c are different, there is a unique $x \in (0,1)$ where I_t assumes its maximum. It is the unique solution on* $[0,1]$ *of the equation*

$$(*) \qquad e^{-K/\dot{f}}(f(x))^{t_2} - (1-f(x))^{t_1} = 0$$

where $K = (c\varepsilon - t_2)H(a) + (t_2 - a\varepsilon)H(c)$ *and* $\varepsilon := t_2 - t_1$.

Proof. At a point $x \in (0,1)$ where I_t takes its maximum, $\dot{I}_t(x) = 0$. Let

$$p(x) = H(f(x)) - x \cdot H(a) - (1-x) \cdot H(c).$$

Notice that $t_1 f(x) + t_2(1 - f(x)) = t_2 - \varepsilon f(x)$. For the sake of readability, we write $\dot{I}_t(x)$ as $\dot{I}_t$, $f(x)$ as f, etc. Then since $\dot{I}_t = 0$,

$$(t_2 - \varepsilon f)\dot{p} - p(-\varepsilon\dot{f}) = 0$$

where

$$\dot{p} = (\ln(1-f) - \ln(f))\dot{f} - H(a) + H(c)$$

Then $(t_2 - \varepsilon f)\dot{p}$ is equal to

$$t_2\dot{f}\ln(1-f) - t_2\dot{f}\ln(f) - t_2H(a) + t_2H(c) - \varepsilon f\dot{f}\ln(1-f) + \varepsilon f\dot{f}\ln(f) + \varepsilon fH(a) - \varepsilon fH(c)$$

and

$$p\varepsilon\dot{f} = -\varepsilon\dot{f}f\ln(f) - \varepsilon\dot{f}(1-f)\ln(1-f) - \varepsilon\dot{f}xH(a) - \varepsilon\dot{f}H(c) + x\varepsilon\dot{f}H(c)$$

When we add these expressions, the first term of $p\varepsilon\dot{f}$ cancels with the sixth term of $(t_2 - \varepsilon f)\dot{p}$, and the second term of $p\varepsilon\dot{f}$ causes the first and fifth terms of $(t_2 - \varepsilon f)\dot{p}$ to vanish leaving $t_1\dot{f}\ln(1-f)$. Thus, our sum reduces to

$$\Big(-t_2\dot{f}\ln(f) - t_2H(a) + t_2H(c) + \varepsilon fH(a) - \varepsilon fH(c)\Big) + \Big(t_1\dot{f}\ln(1-f) - \varepsilon\dot{f}xH(a) - \varepsilon\dot{f}H(c) + x\varepsilon\dot{f}H(c)\Big)$$

which by $f = \dot{f}x + c$ and properties of logarithms is

$$\dot{f}\ln\left(\frac{(1-f)^{t_1}}{f^{t_2}}\right) + (c\varepsilon - t_2)H(a) + (t_2 - a\varepsilon)H(c)$$

Thus, x is a zero of

$$g(x) = e^{-K/\dot{f}}(f(x))^{t_2} - (1 - f(x))^{t_1}$$

Now suppose g had two distinct zeroes. Then its derivative would have to be zero at some point in between. But

$$\dot{g} = \dot{f}\left(e^{-K/\dot{f}}t_2(f(x))^{t_2-1} + t_1(1-f(x))^{t_1-1}\right)$$

which is never zero as the product of $\dot{f} = a - c \neq 0$ and a positive number. □

To reassure ourselves that the equation above is valid, let us consider a few special cases of it. In [1], for the timed Z channel, $c = 0$, so $K = -t_2H(a)$, $f(x) = ax$ and we get

$$e^{t_2H(a)/a}(ax)^{t_2} - (1 - ax)^{t_1} = 0$$

This equation easily follows from the following equation encountered in [1]:

$$\left(\frac{t_2}{a}\right)H(a) = t_1\log\left(\frac{1-f(x)}{f(x)}\right) - \varepsilon\log(f(x))$$

Another case is the untimed case: if $\varepsilon = 0$, then $K = t_2(H(c) - H(a))$, so our equation is now

$$e^{t_2(H(a)-H(c))/(a-c)}(f(x))^{t_2} - (1 - f(x))^{t_1} = 0$$

Taking logs of both sides and simplifying yields

$$\ln\left(\frac{1-f(x)}{f(x)}\right) = \frac{H(a) - H(c)}{a - c}$$

which is the equation we have to solve to calculate the capacity in the untimed case. Notice that the dependence on time is eliminated when $\varepsilon = 0$ regardless of the value of $t_2 = t_1$. Finally, in the case of a *binary symmetric channel*, where the probability of a bit flip is p, we have $a = 1 - p = d$ and $b = p = c$, so our equation takes the form

$$u^{t_2} - (k - u)^{t_1} = 0$$

where $u = f(x)e^{H(p)}$ and $k = e^{H(p)}$.

Theorem 2.3 *Let $x \in (0,1)$ be the solution of $(*)$ for a timing channel with two symbols and $a \neq c$. Then its capacity, measured in bits per unit time, is*

$$\frac{1}{\ln(2)} \cdot \left(\frac{H(a)(c-1) + H(c)(1-a)}{(a-c)} - \ln(f(x)) \right) \cdot \frac{1}{t_1}$$

Proof. Going backward from $(*)$, we see that x satisfies

$$-\frac{K}{t_1 \dot{f}} + \frac{t_2}{t_1} \ln(f(x)) = \ln(1 - f(x)).$$

Substituting this into I_t, abbreviating $f(x)$ as f, we have

$$\begin{aligned}
I_t(x) &= \frac{-f\ln(f) - (1-f)\ln(1-f) - xH(a) - (1-x)H(c)}{t_1 f + (1-f)t_2} \\
&= \frac{-f\ln(f) - (1-f)\left(-\frac{K}{t_1\dot{f}} + \frac{t_2}{t_1}\ln(f)\right) - xH(a) - (1-x)H(c)}{t_2 - \varepsilon f} \\
&= \frac{-t_1\dot{f}f\ln(f) - (1-f)\left(-K + t_2\dot{f}\ln(f)\right) - t_1\dot{f}xH(a) - t_1\dot{f}(1-x)H(c)}{t_1\dot{f}(t_2 - \varepsilon f)}
\end{aligned}$$

Now using $\varepsilon + t_1 = t_2$, $(f - c) = \dot{f}x$, we get

$$\begin{aligned}
I_t(x) &= \frac{-t_1\dot{f}f\ln(f) - (1-f)\left(-K + (\varepsilon + t_1)\dot{f}\ln(f)\right) - t_1(f-c)H(a) - t_1(\dot{f} - (f-c))H(c)}{t_1\dot{f}(t_2 - \varepsilon f)} \\
&= \frac{(1-f)K + \dot{f}\ln(f)(-t_1 f - (1-f)(\varepsilon + t_1)) - t_1(f-c)H(a) - t_1(a-f)H(c)}{t_1\dot{f}(t_2 - \varepsilon f)} \\
&= \frac{(1-f)K + \dot{f}\ln(f)(\varepsilon f - t_2) - t_1(f-c)H(a) - t_1(a-f)H(c)}{t_1\dot{f}(t_2 - \varepsilon f)}
\end{aligned}$$

Now we focus on the expression $(1-f)K - t_1(f-c)H(a) - t_1(a-f)H(c)$. It equals

$$H(a)((1-f)(c\varepsilon - t_2) - t_1(f-c)) + H(c)((1-f)(t_2 - a\varepsilon) - t_1(a-f))$$

which is

$$H(a)(c-1)(t_2 - \varepsilon f) + H(c)(1-a)(t_2 - \varepsilon f)$$

Putting everything together, we get

$$\begin{aligned}
I_t(x) &= \frac{(t_2 - \varepsilon f)(H(a)(c-1) + H(c)(1-a) - \dot{f}\ln(f))}{t_1\dot{f}(t_2 - \varepsilon f)} \\
&= \frac{H(a)(c-1) + H(c)(1-a)}{(a-c)t_1} - \frac{\ln(f)}{t_1}
\end{aligned}$$

Finally, because capacity is measured in bits, we convert our logarithms to base 2 by multiplying by $1/\ln(2)$. □

Given that the capacity calculation depends entirely on our ability to compute the solution of $(*)$, we now turn to methods for calculating it which are provably correct.

3 Algorithms for calculating the capacity

We now consider methods for calculating the unique solution of $g(x) = 0$. First notice that it is enough to solve the equation $h(u) = 0$ where

$$h(u) = e^{-K/\dot{f}} u^{t_2} - (1-u)^{t_1}$$

and then obtain $x = f^{-1}(u)$. One way to solve $h(u) = 0$ is to use the bisection method since h changes sign on $[0,1]$ (i.e. $h(0) = -1 < 0$ and $h(1) > 0$). Here is a one point method.

Theorem 3.1 *Let $\phi : [0,1] \longrightarrow \mathbb{R}$ be the map*

$$\phi(x) = x - \frac{h(x)}{M}$$

where the constant M is given by

$$M = t_2 e^{-K/\dot{f}} + t_1$$

For any $x \in [0,1]$, the sequence $(\phi^n(x))$ converges to the unique zero r of h on $[0,1]$.

Proof. First, we claim that $\dot{\phi}(x) \in (0,1)$ for all $x \in [0,1]$. Since $\dot{\phi} = 1 - \dot{h}/M$, all we have to show is that $0 < \dot{h}(u) < M$ for $u \in [0,1]$. First

$$\dot{h}(u) = e^{-K/\dot{f}} t_2 u^{t_2} + t_1(1-u)^{t_1 - 1} > 0$$

if $0 < u \leq 1$ while $\dot{h}(u) = t_1 > 0$ for $u = 0$. So $\dot{h} > 0$ on $[0,1]$. Next we see

$$\dot{h}(0) = t_1, \quad \dot{h}(u) < M \text{ for } u \in (0,1), \quad \dot{h}(1) < t_2 e^{-K/\dot{f}}$$

so that $\dot{h} < M$ on $[0,1]$. Let

$$c_\phi := \sup_{x \in [0,1]} \dot{\phi}(x)$$

By the continuity of $\dot{\phi}$, there is $z \in [0,1]$ such that $\dot{\phi}(z) = c_\phi$ and so $0 \leq c_\phi < 1$. Now if we are given distinct points $x < y \in [0,1]$, then by the mean value theorem, there is some $p \in (x,y)$ with

$$|\phi(x) - \phi(y)| = |\dot{\phi}(p)| \cdot |x - y| = \dot{\phi}(p)|x - y|$$

which means $|\phi(x) - \phi(y)| \leq c_\phi |x - y|$. If r is the unique zero of h and $x \in [0,1]$ is any other point,

$$|\phi^n(x) - r| \leq c_\phi^n |x - r|$$

which implies $\phi^n(x) \longrightarrow r$ since $c_\phi < 1$. □

The bisection is a "bracketing method" – one of its advantages is that we always have some idea of how close we are to the zero since we carry both an upper and a lower bound, while a potential advantage of the one point method above is that a smaller upper bound M on $\dot{h}$ will improve its convergence (though useful estimates of its rate of convergence are unknown to us). Our next method originates from [2]. It attempts to combine the advantages of both the bisection and the one point method in the last theorem.

Definition 3.2 Let $\mathbf{IR} = \{[a,b] : a \leq b \ \& \ a, b \in \mathbb{R}\}$ ordered by reverse inclusion $\sqsubseteq$. We use the following operators on $\mathbf{IR}$:

- $l : \mathbf{IR} \to \mathbb{R} :: [a,b] \mapsto a$
- $m : \mathbf{IR} \to \mathbb{R} :: [a,b] \mapsto (a+b)/2$
- $r : \mathbf{IR} \to \mathbb{R} :: [a,b] \mapsto b$

These are abbreviated $l_x := l(x)$, $r_x := r(x)$ and $m_x := m(x)$. Let

$$\Box h := \{x \in \mathbf{IR} : [0,1] \sqsubseteq x \sqsubseteq [r]\}$$

denote the set of intervals where h changes sign.

For instance, the bisection method defines a function $\text{split}_h : \Box h \to \Box h$ given by

$$\text{split}_h(x) = \begin{cases} [l_x, m_x] & \text{if } h(m_x) > 0; \\ [m_x, r_x] & \text{otherwise.} \end{cases}$$

If we iterate this function beginning, say, from $x = [0,1]$, then $(\text{split}_h^n(x))$ is a decreasing sequence of intervals which contain r and whose lengths tend to zero.

Theorem 3.3 *Iterating the mapping $s_h : \Box h \to \Box h$ given by*

$$s_h(x) = \begin{cases} [l_x, m_x - (h(m_x)/M)] & \textit{if } h(m_x) > 0; \\ \\ [m_x + (|h(m_x)|/M)\,, r_x] & \textit{otherwise;} \end{cases}$$

is an algorithm for calculating r. That is, it produces a decreasing sequence of intervals which contain r and whose lengths tend to zero:

$$\bigsqcup_{n \geq 0} s_h^n(x) = [r],$$

for all $x \in \Box h$. Thus, for all $x \in \Box h$, if $x_n \in s_h^n(x)$ for each n, then $x_n \to r$.

This result is a special case of a more general result given in [2]. Its proof uses the fact that h has a unique zero r with $h > 0$ to the right of r and $h < 0$ to the left of r. By incorporating an upper bound of h, this method outdoes the bisection method *at every iteration.* While this method does use information about the first derivative in the computation (the upper bound M), it does not use the derivative itself as part of the computation. For this reason, the general version of the algorithm in [2], which applies to the class of Hölder continuous functions, is capable of beating the bisection at every iteration without requiring any differentiability. The nowhere differentiable function introduced in 1872 by Weierstrass is a well-known example of this type.

References

[1] I. S. Moskowitz, S. Greenwald and M. H. Kang. *An analysis of the timed Z-channel.* IEEE Transactions on Information Theory, Vol. 44, No. 7, November 1998.

[2] K. Martin. *B-sides.* Oxford University Computing Lab Research Report, January 2003.